

Text2Sign: Towards Sign Language Production Using Neural Machine Translation and Generative Adversarial Networks

Year: 2020 | Citations: 243 | Authors: Stephanie Stoll, N. C. Camgoz, Simon Hadfield, R. Bowden

Abstract

We present a novel approach to automatic Sign Language Production using recent developments in Neural Machine Translation (NMT), Generative Adversarial Networks, and motion generation. Our system is capable of producing sign videos from spoken language sentences. Contrary to current approaches that are dependent on heavily annotated data, our approach requires minimal gloss and skeletal level annotations for training. We achieve this by breaking down the task into dedicated sub-processes. We first translate spoken language sentences into sign pose sequences by combining an NMT network with a Motion Graph. The resulting pose information is then used to condition a generative model that produces photo realistic sign language video sequences. This is the first approach to continuous sign video generation that does not use a classical graphical avatar. We evaluate the translation abilities of our approach on the PHOENIX14T Sign Language Translation dataset. We set a baseline for text-to-gloss translation, reporting a BLEU-4 score of 16.34/15.26 on dev/test sets. We further demonstrate the video generation capabilities of our approach for both multi-signer and high-definition settings qualitatively and quantitatively using broadcast quality assessment metrics.